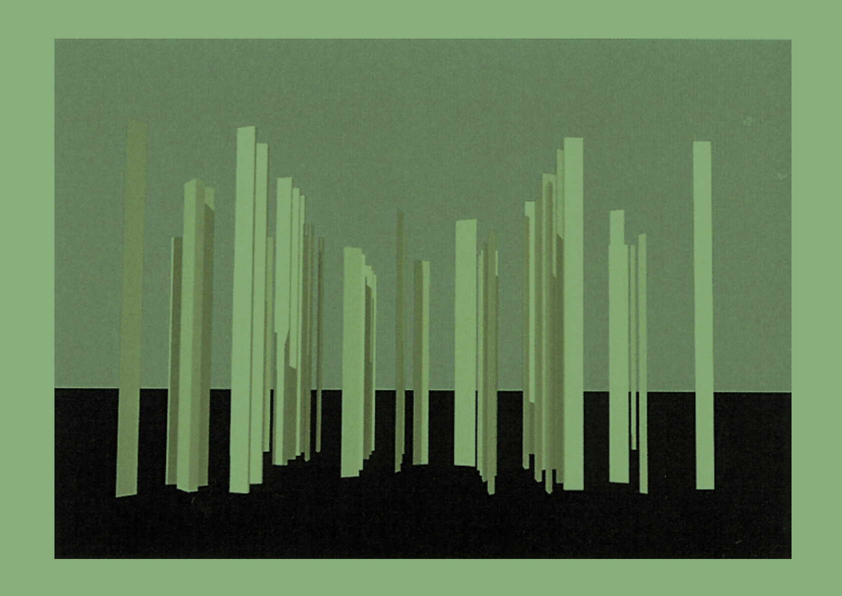
INTRODUCCIÓN A LA ESTADÍSTICA

INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

por

JUANA Mª SÁNCHEZ GONZÁLEZ



CUADERNOS
DEL INSTITUTO
JUAN DE HERRERA
DE LA *ESCUELA DE*
ARQUITECTURA
DE MADRID

3-83-01

INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

por

JUANA Mª SÁNCHEZ GONZÁLEZ

CUADERNOS
DEL INSTITUTO
JUAN DE HERRERA
DE LA *ESCUELA DE*
ARQUITECTURA
DE MADRID

3-83-01

C U A D E R N O S
D E L I N S T I T U T O
J U A N D E H E R R E R A

NUMERACIÓN

2 Área
51 Autor
09 Ordinal de cuaderno (del autor)

TEMAS

1 ESTRUCTURAS
2 CONSTRUCCIÓN
3 FÍSICA Y MATEMÁTICAS
4 TEORÍA
5 GEOMETRÍA Y DIBUJO
6 PROYECTOS
7 URBANISMO
8 RESTAURACIÓN
0 VARIOS

Introducción a la estadística.
Introducción a la estadística descriptiva.

Instituto Juan de Herrera.
Escuela Técnica Superior de Arquitectura de Madrid.
Gestión y portada: Almudena Gil Sancho.
CUADERNO 322.01 / 3-83-01
ISBN-13 (obra completa): 978-84-9728-390-8
ISBN-13: 978-84-9728-391-5
Depósito Legal: M-1920-2012

La serie de cuadernillos dedicados a distintos aspectos de la Estadística que iniciamos con este primero, Introducción a la Estadística Descriptiva, surge como resultado de la experiencia obtenida durante el pasado curso académico de 2010-2011 en el que se implanta el Primer Curso del llamado Plan 2010 en la Escuela Técnica Superior de Arquitectura de Madrid.

En el contenido docente de ese primer curso se incluyen una serie de Talleres en los que se estudian temas, relacionados con las asignaturas que conforman el contenido de la carrera desde un punto de vista más práctico, más directo y manteniendo una relación constante con un alumnado al que se organiza en grupos más reducidos que los dedicados a otras asignaturas.

El Departamento de Matemática Aplicada a la Edificación, al Medio Ambiente y al Urbanismo ha ofertado, entre ellos, el taller de Informática Aplicada. En él se desarrollan distintos Programas Informáticos siempre relacionados con lo que será la actividad profesional del futuro arquitecto.

Uno de los módulos del taller de Informática Aplicada se dedica al conocimiento, manejo y aplicación de programas informáticos que hagan más fácil y posible la elaboración de estudios estadísticos ya que, en nuestros días, es imposible imaginar una investigación relacionada con cualquier tipo de ciencia y donde sea necesario el manejo de una gran cantidad de datos, en la que no se haga uso de la estadística, del ordenador y de programas dedicados expresamente este fin.

Dado que los alumnos que acceden a los estudios de Arquitectura no han cursado, en general, en su opción de Bachillerato, el tipo de Matemática que contempla los conceptos necesarios para acometer con soltura estas prácticas, hemos considerado la posibilidad de ofrecerles un complemento teórico y práctico que les resulte útil y les acerque a este tipo de conocimientos.

Quiero hacer llegar mi agradecimiento a todos los profesores que con sus sugerencias y críticas han ayudado a que este cuaderno se haga realidad, así como a la alumna B. M. Francelin, autora de la imagen de la portada.

Índice

0 Introducción

La Estadística es una rama de las Matemáticas que se ocupa de recoger información sobre muchos y muy diversos temas, de analizarla y sistematizarla con el fin de intentar, posteriormente, predecir hechos o situaciones relacionados con dichos temas.

La palabra Estadística se viene utilizando, principalmente, con dos significados distintos:

a) Como representativa de un conjunto de datos numéricos que, después de ser obtenidos de una forma sistemática y ordenados según un determinado criterio, llegan a formar un conjunto coherente.

b) Dando nombre a una ciencia. En este caso, la Estadística usará un lenguaje específico y tratará, como otra ciencia cualquiera, de buscar las características generales de un determinado colectivo anteponiéndolas a las particulares de cada uno de los elementos que lo componen. Posteriormente, el estudio, sistemático y ordenado de esas características generales, servirá para, además de describir el fenómeno y, en consecuencia, conseguir comprenderlo mejor y/o poder hacer predicciones sobre su futuro comportamiento.

1 Conceptos generales

1.1 Definición de términos estadísticos

Cualquier estudio estadístico tiene como objeto un conjunto de elementos de cualquier naturaleza o condición.

Al conjunto en sí se le llamará *población*. A cada uno de sus componentes, *elemento* o *individuo* y al número de los que componen la población, *tamaño* de la población. El tamaño podrá ser finito o infinito.

Una *muestra* será una parte cualquiera de la población.

Una vez elegida la población o, en su caso, la muestra, se llama *encuesta* la recogida de información correspondiente a una muestra. Cuando la muestra está compuesta por toda la población, la recogida de información se denomina *censo*.

1.2 Caracteres estadísticos

Las características o propiedades a estudiar de una población o de una muestra se pueden agrupar en dos tipos: propiedades *cualitativas* y propiedades *cuantitativas*.

a) Las propiedades *cualitativas*, llamadas también *atributos*, se definen mediante “palabras” y responderían a situaciones como podrían ser los estudios relativos al estado civil de los habitantes de un pueblo, el color de pelo de los alumnos de una clase....

b) Las propiedades *cuantitativas* son las que pueden valorarse con números y, a su vez, podrán ser *discretas* o *continuas*.

   Los datos obtenidos del estudio de las variables *discretas* se pueden cuantificar con números enteros, caso del número de hijos de un determinado grupo de matrimonios, número de lápices que llevan en la cartera los alumnos de una clase....

   Las variables *continuas*, por su naturaleza, se podrán evaluar de forma más precisa con números decimales. Sería el caso, por ejemplo, del conjunto de datos recogidos tras el análisis del peso de la plantilla de jugadores de un determinado equipo de fútbol, las estaturas de los soldados de una compañía, las notas medias obtenidas por un grupo de 1º de Arquitectura en el examen final de la asignatura de Geometría.....

1.3 El método estadístico

En cualquier caso, tengamos población o muestra, variables cualitativas o cuantitativas, en la elaboración de los estudios estadísticos es necesario llevar a cabo una serie de etapas a saber:

a) La recogida de datos sin elaborar (es lo que se llama *observación* de la población o muestra).

b) La ordenación y presentación de los mismos. Tanto la ordenación como la presentación de los datos recogidos en una determinada observación recurre, en su sistematización, a *tablas* y *gráficos*.

c) La elaboración del resumen de la información que proporcionan las tablas y gráficos obtenidos anteriormente.

d) La de extraer información sobre las características analizadas. Este es quizá el aspecto más importante de todos.

En un análisis estadístico cualquiera, de los tres primeros apartados se ocupa la llamada **Estadística Descriptiva.** Del cuarto, la **Inferencia Estadística**. Para su manejo se utiliza, con bastante frecuencia, el **Cálculo de Probabilidades**.

2 Estadística descriptiva

Dentro de la Estadística, la Descriptiva, es una parte de esta ciencia que se dedica a recoger y analizar datos de una determinada propiedad, elegida entre las que puede tener un grupo de individuos, de forma que, una vez sistematizados en tablas, representados en gráficos y tras haber calculado una serie de parámetros, sea posible ver cómo se comporta.

2.1 Tipos de análisis de variables estadísticas

El análisis de una población o muestra será *univariado* si se considera el estudio de sólo una de sus características o propiedades. Será *multivariado* si se analizan dos o más de ellas.

2.1.1 Análisis univariado

Como se indicó con anterioridad, en cualquier análisis estadístico, en este caso de sólo una propiedad de la muestra o la población elegida, se necesita una serie de etapas:

a) El trabajo de campo, que recoge los valores o modalidades que toma la propiedad analizada.

b) La elección de la mejor forma de ordenarlos en tablas y de presentarlos en gráficos ya que, de ser acertada, puede empezar a aclarar conceptos.

c) La elaboración del resumen de la información que proporcionan las tablas y gráficos citados.

d) La obtención de información, sobre la característica estudiada, tras el análisis de ese resumen.

En la práctica, la forma de observar una sola variable depende de los siguientes factores:

a) Del número de datos obtenido.

b) Del número de valores distintos que tome la variable en cada caso analizado.

Teniendo en cuenta estos dos factores tendríamos tres tipos distintos de estudios estadísticos de una sola variable:

I) Estudios que constan de pocas observaciones.

II) Estudios con muchas observaciones, pero con pocos valores distintos.

III) Estudios con muchas observaciones y muchos valores distintos.

En el primer caso, la ordenación de los valores no ofrece dificultad.

En el segundo sería fundamental tener en cuenta la *frecuencia* o número de veces que cada valor aparece repetido.

En el tercero convendría agrupar los valores en distintos *intervalos* o *clases* para, posteriormente, determinar la frecuencia de cada uno de ellos.

2.1.2 Tablas estadísticas

La primera acción a llevar a cabo, después de recoger una serie de datos encaminados a elaborar un determinado estudio estadístico, es la de ordenarlos en una serie de *tablas* y *gráficos* que permitan ver de una forma más práctica y directa las tendencias que presentan y las reglas que siguen.

Las elaboración de tablas estadísticas constituye uno de los pasos más importante dentro de un estudio estadístico, ya que son las encargadas de proporcionar toda la información global que se necesita para plantear o conseguir los objetivos que se persiguen en dicho estudio.

Sirva como ejemplo el siguiente gráfico que representa una tabla en la que se han agrupado, en columnas, los resultados observados, en una población de 22 individuos, de una propiedad cualquiera que puede tomar los valores cuantitativos discretos de 5, 6, 7 y 8. Como resultado se ha llegado a la conclusión de que en 3 individuos se repite el valor 5, en 7 el valor 6 y en 8 de ellos el valor 8 (Tabla 1).

En la primera columna se han representado los valores que ha tomado la variable analizada.

En la segunda, el número de individuos que toman ese valor. Es lo que se llama *frecuencia absoluta*.

La tercera columna representa la *frecuencia relativa*, es decir, el resultado de dividir cada frecuencia absoluta por el tamaño de la población. Por supuesto, la suma de las frecuencias absolutas nos dará como resultado la población y la suma de las relativas, la unidad.

Existen muchas formas distintas de elaborar tablas de frecuencias. Con el uso del ordenador esta tarea se ha facilitado considerablemente.

Es necesario aclarar, que el total anotado en la tabla de frecuencias relativas se habrá de redondear, ya que tiene que representar al total de la población. En el cálculo llevado a cabo en un caso real, es posible que el valor resultante de sumar los valores, dependa, en gran parte, del número de decimales utilizado en el trabajo y no coincida exactamente con la unidad

Datos	Frecuencia absoluta	Frecuencia relativa
5	3	0,14
6	7	0,32
7	4	0,18
8	8	0,36
	Total 22	Total 1,00

Tabla 1. Tabla de frecuencias

A partir de unas tablas ya elaboradas se pueden conseguir con más facilidad tanto gráficos como determinadas cantidades, conocidas como *parámetros estadísticos*, que pueden ayudar a describir un determinado aspecto del proceso. Los gráficos más utilizados y los parámetros estadísticos se analizarán posteriormente.

3 Representaciones gráficas

Una vez realizada la observación de una variable y obtenido una serie de datos, el ordenarlos en distintos tipos de tablas para que puedan ser utilizados en el dibujo de determinados gráficos facilita su manejo. Aspectos importantes del estudio que se esté llevando a cabo pueden no ser tan evidentes manejando sólo una tabla como observando figuras que los representen.

La elección de las representaciones gráficas vendrá en función del tipo de propiedades que se quieran estudiar, y cambiará según se trate de variable cuantificable (discreta o continua) o atributo.

En el primer caso, el de variable discreta, las representaciones más adecuadas, y usadas generalmente con distintos aspectos estéticos, son la del *diagrama de barras* y el *polígono de frecuencias* que se le puede asociar. Para las variables continuas la representación más explícita la constituye el *histograma.*

El atributo, la variable cualitativa, se viene representando con eficacia por medio de distintos tipos de diagramas de aspecto circular; útiles, también, en la representación de datos de carácter cuantitativo.

3.1 Diagrama de barras

En este tipo de gráfico los datos obtenidos se representan sobre un sistema de ejes coordenados en los que en uno, cualquiera de ellos, se recogen los valores obtenidos de la observación del conjunto problema. En el eje restante se representan las frecuencias con que se repiten cada uno de los valores considerados. Pueden usarse frecuencias absolutas o relativas.

Las barras pueden representarse en horizontal o en vertical. En los dos casos la altura de la barra se tomaría proporcional a la frecuencia de los valores obtenidos y paralela al eje que la representase.

Existen distintas formas de dibujar este tipo de gráficos. A continuación se plantean algunos ejemplos.

Ejemplo 1.

Se ha preguntado a una serie de personas por qué compran los regalos “on line” durante las temporadas festivas. Las razones que dieron se categorizaron en:

a) La facilidad de comparar productos.

b) La conveniencia.
c) Los gastos de envío gratis.
d) La variedad en la selección.
e) La velocidad de entrega.

Se pide llevar los datos a una tabla (Tabla 2) con el fin de dibujar un diagrama de barras (Fig. 1) representativo de la situación.

Razón	Porcentaje (%)
Comparación de compra	23
Conveniencia	33
Gastos de envío gratis	34
Variedad de la selección	6
Velocidad	4
Total	100

Tabla 2

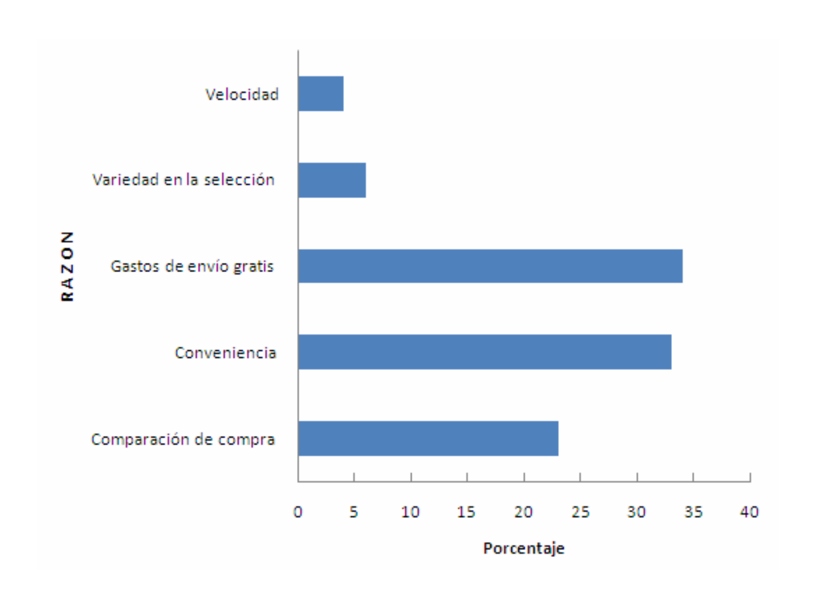


Figura 1. Diagrama de barras horizontales correspondiente a la tabla 2

Los mismos datos nos habrían servido para dibujar un diagrama de barras verticales (Fig. 2).

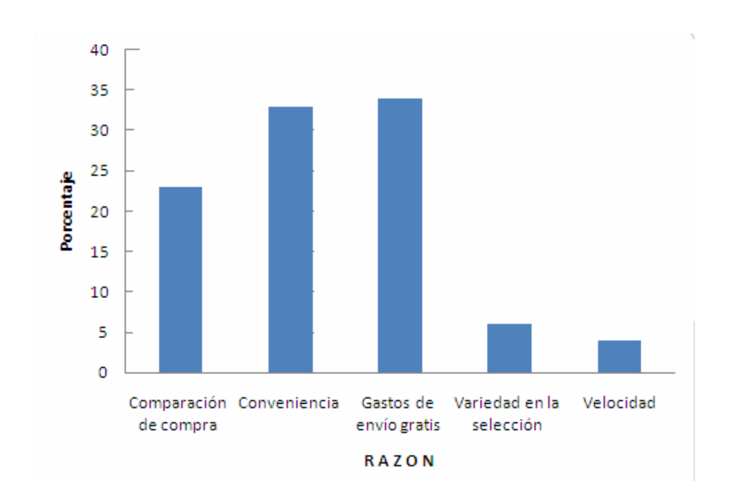


Figura 2. Diagrama de barras verticales correspondiente a la tabla 2

Ejemplo 2.

En este otro caso (Fig. 3) se van a representar los valores numéricos obtenidos en el análisis de un caso de variable cuantitativa en un diagrama de barras verticales.

Supongamos que en un determinado colegio se les ha preguntado a un conjunto de alumnos la nota que han sacado en un examen de Matemáticas. Las notas se han movido entre el 0 y el 7.

A la pregunta de cuál ha sido su nota, dos alumnos han respondido que 0. Tres de ellos que 1. Cinco han obtenido un 2. Las notas de 3 y 4 las han conseguido el mismo número de alumnos, ocho. Sólo tres han obtenido un 5. Dos un 6 y la máxima nota de 7, uno de ellos.

Al dibujar los resultados en un diagrama de barras, en el eje horizontal se han representado los valores que se han tomado de la variable, la nota obtenida en un determinado examen por un grupo de alumnos, y en el eje vertical el número de veces, la frecuencia, f_i, con la que se repite ese valor en el estudio.

Para mayor claridad en el dibujo, el valor 0 del conjunto de valores no se ha colocado en el origen del sistema de ejes coordenados.

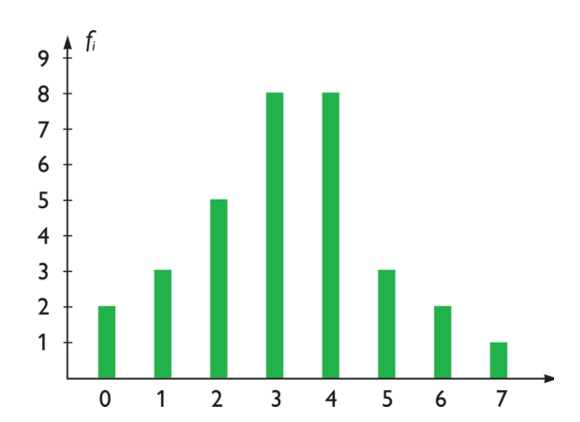


Figura 3. Diagrama de barras verticales correspondiente al ejemplo 2

Ejemplo 3.

Los diagramas de barras pueden agrupar en el mismo gráfico observaciones de variables distintas con el fin de comparar resultados (Fig. 4).

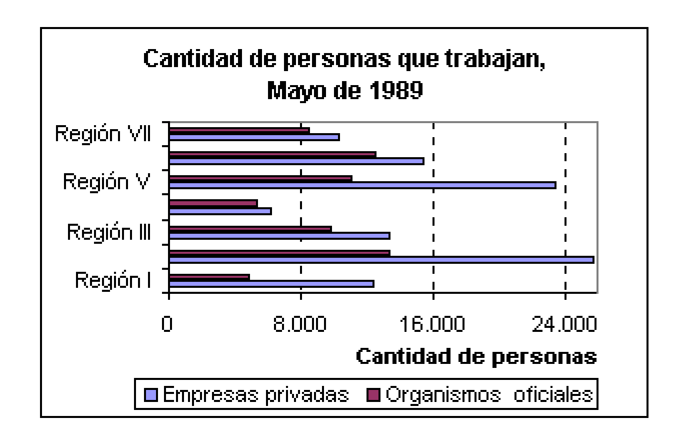


Figura 4. Diagrama de barras horizontales

La interpretación del gráfico nos indica cómo en las distintas regiones en las que se han recabado los datos necesarios para elaborar el gráfico, el empleo es mucho más alto, en general, en las empresas privadas. En algunas ocasiones llega a doblar el de los organismos oficiales.

3.2 Polígono de frecuencias

En un diagrama de barras cualquiera, si unimos, de forma consecutiva, los puntos medios de la parte superior de cada una de las barras con un segmento, se obtiene una línea quebrada que se denomina polígono de frecuencias (Fig. 5).

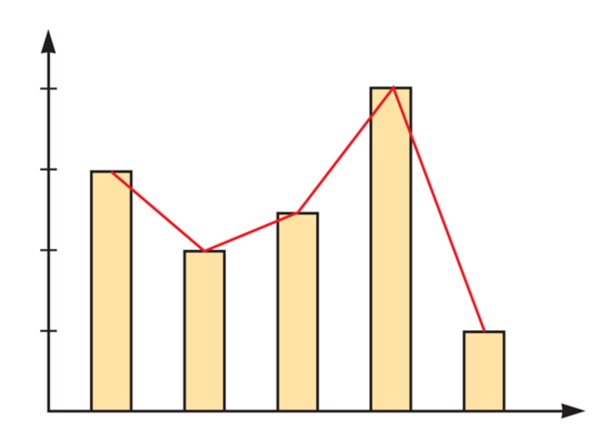


Figura 5. Ejemplo de polígono de frecuencias asociado a un diagrama de barras

En ocasiones puede, incluso, sustituir al diagrama de barras utilizado para obtenerlo, prescindiendo del mismo (Fig. 6).

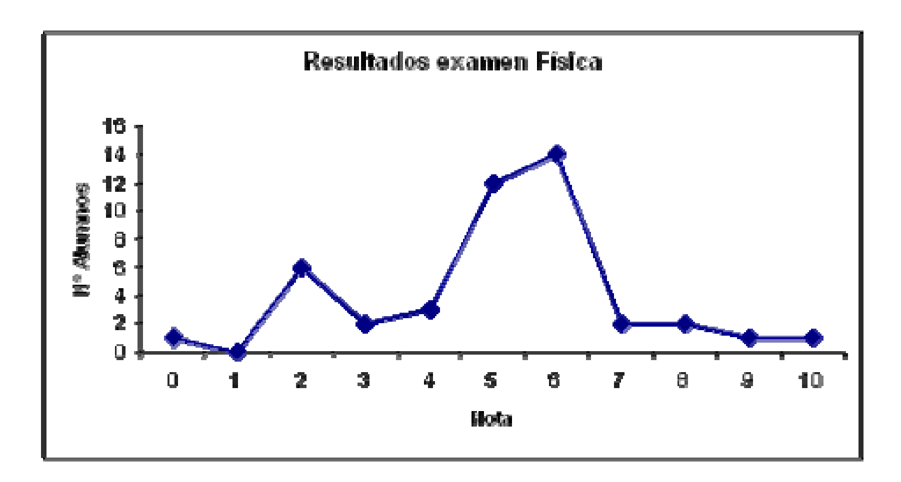


Figura 6. Polígono de frecuencias

El polígono de frecuencias, obtenido de un diagrama de barras que no se dibuja, representa el número de alumnos que han obtenido una determinada nota en un examen de Física que se ha calificado del 0 al 10.

3.3 Pictogramas

En ocasiones, para que el gráfico sea más representativo del tipo de variable analizada, las barras pueden sustituirse por figuras alusivas a la naturaleza del estudio (Fig. 7).

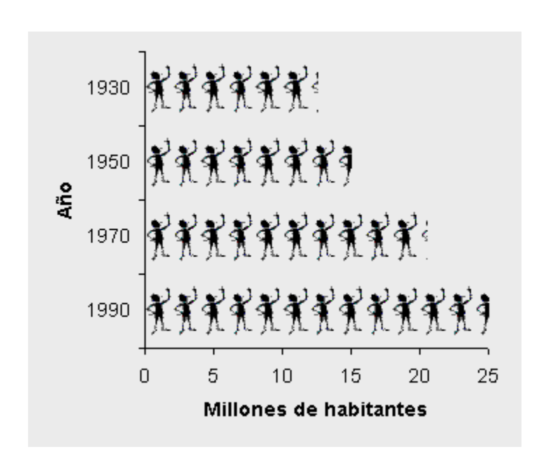


Figura 7. Pictograma de barras horizontales

En la imagen precedente se representa en el eje horizontal la población de un determinado lugar (en millones de habitantes), y en el eje vertical las fechas sucesivas (en años) en las que se ha llevado a cabo el estudio, ordenadas de mayor a menor.

En este caso tampoco se ha tomado el 0 del eje vertical en el origen de coordenadas. Se considera sólo el intervalo temporal estudiado. El resultado es equivalente a un diagrama de barras horizontales.

3.4 Histogramas

El histograma es un gráfico estadístico que también utiliza un sistema de ejes coordenados. En el eje horizontal se representan los valores que puede tomar la variable agrupados en intervalos definidos previamente con una amplitud determinada. En el eje vertical, las frecuencias absolutas o las densidades de frecuencia (valor resultante de dividir la frecuencia absoluta por la amplitud del intervalo correspondiente). Es muy adecuado para variables cuantitativas continuas capaces de tomar valores decimales como pesos, estaturas.... El histograma se basa en el dibujo de rectángulos de base la amplitud del intervalo y de altura o área proporcional a la frecuencia.

Ejemplo 4.

La tabla 3 recoge los datos obtenidos durante el estudio de una variable cuantitativa continua: la estatura de una muestra de personas. Para sistematizarlo se agrupan las edades en intervalos de 10 centímetros de amplitud. Se comienza con una estatura de 155 centímetros hasta llegar a una de 195 centímetros de manera que, en los intervalos, estén incluidos todos los elementos de la muestra analizada.

Las columnas representan sucesivamente los intervalos elegidos, la frecuencia absoluta, la frecuencia relativa y el porcentaje de indivíduos incluidos en cada intervalo.

Intervalos	Frecuencia absoluta	Frecuencia relativa	%
[155-165)	6	0,30	30
[165-175)	10	0,50	50
[175-185)	3	0,15	15
[185-195)	1	0,05	5
Total	20	1,00	100

Tabla 3. Tabla de frecuencias del ejemplo 4

Para dibujar el histograma correspondiente se utilizarían las amplitudes de los intervalos como base de los rectángulos, y como altura las frecuencias absolutas o las densidades de frecuencia (resultado de dividir cada frecuencia absoluta por la amplitud de su intervalo correspondiente). Así cada rectángulo representaría, con su altura o con su superficie, al número de individuos incluidos en cada intervalo.

En ocasiones el dato facilitado para interpretar los resultados de un estudio es, precisamente, su representación gráfica. En la figura 8 se facilita el dibujo de un histograma que representa la energía eléctrica que han consumido una serie de empresas durante el mes de julio.

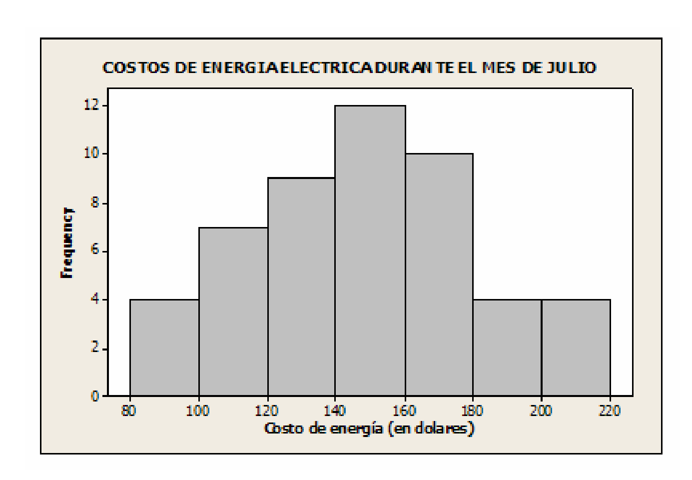


Figura 8. Histograma

Un ejercicio inverso al presentado en el ejemplo 4 podría consistir, a la vista del gráfico anterior (Fig. 8), en plantear, para su posterior análisis, cuestiones como las siguientes:

a) Construir la tabla de frecuencias correspondiente.
b) Construir un diagrama circular que lo represente.
c) Redactar las conclusiones pertinentes.

También en un histograma se podría dibujar su polígono de frecuencias. Bastaría unir con líneas, de forma consecutiva, los puntos medios de los rectángulos dibujados.

3.5 Representación gráfica de variables cualitativas (atributos)

Cuando se elige un determinado diagrama estadístico, lo que se persigue es conseguir que proporcione la representación más clara del suceso que se estudia, evitando aquellos detalles que puedan producir una representación errónea.

En los casos anteriores, el hecho de que los resultados de los fenómenos estudiados se pudiesen expresar con números, aconsejaba la utilización de diagramas geométricos claros.

En el caso de variables en las que los posibles resultados se clasifiquen en distintas modalidades, los sistemas gráficos a utilizar son muy variados. Los más usuales son los *diagramas de sectores* o *diagramas circulares.*

3. 5. 1 Diagramas circulares

Son representaciones que tienen como base círculos en los que se definen sectores circulares de superficie proporcional a la frecuencia, en general, del dato representado.

En la figura 9, por ejemplo, se representa el porcentaje de superficie que se dedica, en una determinada población a los distintos tipos de equipamientos de la total dedicada a dotaciones.

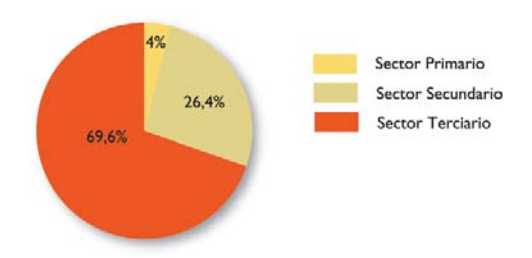


Figura 9. Gráfico de sectores

Los sectores se presentan en colores distintos, o se rayan de formas diferentes, para que la representación sea más explícita.

Ejemplo 6.

A un grupo de 30 alumnos de 4º de la ESO se les ha preguntado si practican algún deporte El resultado de la encuesta se ha recogido en la tabla 4. En ella se ha reflejado el ángulo central de la circunferencia correspondiente a cada uno de los sectores circulares que representarán, con su superficie, cada una de las categorías resultantes de la observación.

Deporte	Alumnos	Ángulo
Baloncesto	12	144º
Natación	3	36
Fútbol	9	108º
Sin deporte	6	72
Total	30	360º

Tabla 4

El diagrama circular representativo de esta situación sería el de la figura 10.



Figura 10. Diagrama de sectores correspondiente a la tabla 4

3. 5. 2 Pirámides de Población

En los apartados anteriores se han propuesto ejemplos de representaciones gráficas que tenían en cuenta el tipo de variable analizada: variables cuantitativas y cualitativas. Sin embargo, además, existen unas representaciones gráficas, muy utilizadas en estadística, que permiten unir información de valores cuantitativos y cualitativos simultáneamente: las pirámides de población. Son una combinación de diagramas de barras e histogramas.

Existen diferentes tipos de pirámides de población de forma que, sólo con observar su apariencia, ya pueden proporcionar datos sobre el problema planteado (Fig. 11).

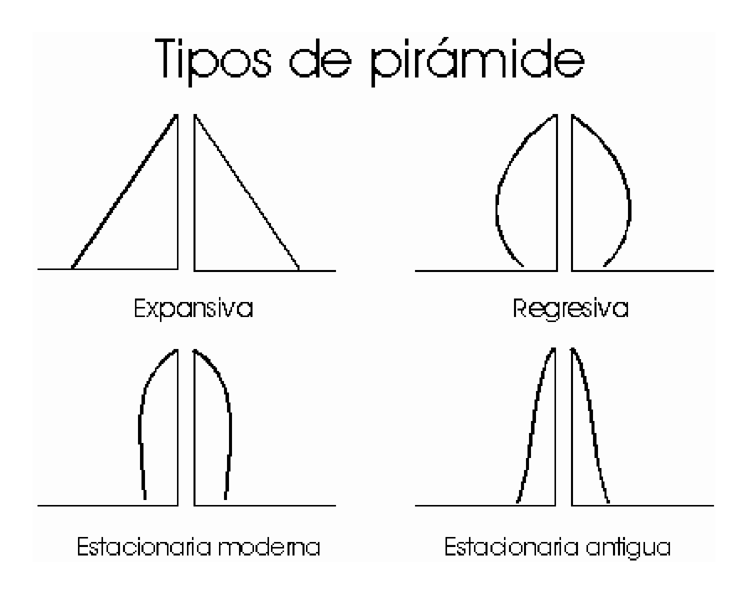


Figura 11

La *expansiva*, por ejemplo, estaría presente en estudios de colectivos pertenecientes al Tercer Mundo en los que exista una gran tasa de natalidad y la esperanza de vida no sea muy alta. La *regresiva* sería representativa de países desarrollados en las que la natalidad ha descendido considerablemente y la esperanza de vida sube con el tiempo. Las *estacionarias* de cualquier tipo, por su parte, pertenecerían a modelos de población en los que existe una relativa igualdad entre las generaciones jóvenes y adultas.

El caso representado en la figura 12, nos da información sobre una población de un país desarrollado en el que se ha llevado a cabo un estudio en el que están presentes dos tipos de variables de una misma población: la variable cualitativa, hombre/mujer, y la cuantitativa, edad, analizada por intervalos.

Para elaborar el gráfico se han considerado histogramas horizontales, con intervalos de 4 años de amplitud como base, para clasificar al conjunto de la población desde su nacimiento hasta la edad de más de 85 años en ambos casos.

En el eje horizontal se representa la población expresada en cualquier unidad, en miles de personas por ejemplo, y en el eje vertical, enfrentando hombres y mujeres, los intervalos de edades.

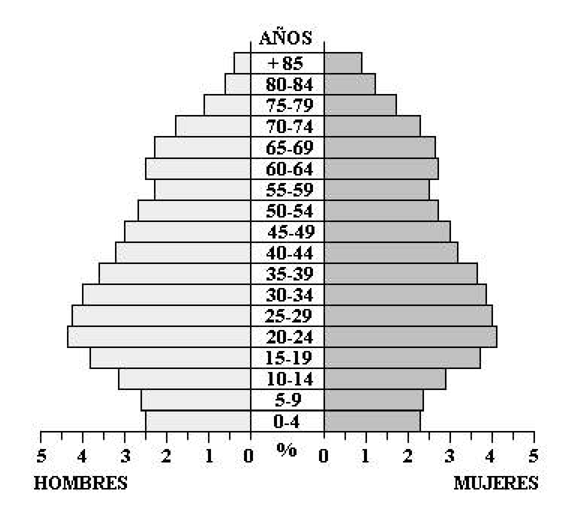


Figura 12. Pirámide de edad

Como se puede observar, el mayor número de hombres tienen una edad comprendida entre los 20 y 24 años. En el caso de las mujeres esa circunstancia se da a la misma edad. Es significativo que, naciendo más hombres que mujeres, sea más del doble el número de mujeres, respecto al de hombres, que superan los 85 años.

4 Parámetros estadísticos

Una vez planteado el estudio y observada la población o muestra elegida en cada caso, analizados y representados los datos en las tablas o formas gráficas que se consideren más adecuadas, se podrán sacar consecuencias de la propiedad, o propiedades que se hayan analizado, mediante la utilización de los parámetros estadísticos: valores numéricos adicionales que ayudarán a completar y sistematizar la información obtenida.

4.1 Parámetros estadísticos de posición

También llamados centrales o promedios son la *Media*, la *Moda* y la *Mediana*. Representan las medidas más comunes en este tipo de estudios y se les califica como valores de posición porque responden, especialmente, a la forma en que se reparten las frecuencias de los datos obtenidos de la población o muestra analizada.

4. 1. 1 Media

Para definir la media vamos a suponer que estamos llevando a cabo el estudio de un suceso S y que de su observación se han obtenido una serie de datos, x_i :

$$x_1, x_2, x_3, ..., x_n$$

Definimos como media aritmética, $\overline{x}$, al valor que se obtiene al dividir la suma de todos los datos, x_i , por el número de ellos, n. Es decir:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

La media, $\overline{x}$, se medirá en las mismas unidades que x_i .

Si, como es frecuente, los datos de la nuestra se repiten, y si llamamos n_i a la frecuencia absoluta del dato (número de veces que se repite ese valor x_i en el estudio), entonces la media se escribe:

$$\overline{x} = \frac{x_1 n_1 + x_2 n_2 + ... + x_m n_m}{N} = \frac{1}{N}\sum_{i=1}^{n} x_i n_i$$

Donde

$$N = n_1 + n_2 + ... + n_m$$

En esta expresión se considera que todos los datos obtenidos tienen la misma importancia o fiabilidad. No obstante, puede darse el caso de que, según el criterio adoptado al hacer el estudio, se de más importancia a unos datos que a otros. En este caso la media se llama *ponderada* y su expresión sería la siguiente:

$$\overline{x}_p = \frac{x_1 n_1 a_1 + x_2 n_2 a_2 + ... + x_m n_m a_m}{N_p} = \frac{1}{N_p} \sum_{i=1}^{n} x_i n_i a_i$$

Siendo a_i son las distintas ponderaciones o pesos que se adjudican a los datos y

$$N_p = n_1 a_1 + n_2 a_2 + ... + n_m a_m$$

4. 1. 2 Moda

Al hacer un estudio de una población, o de una muestra de esa población, puede suceder que unos valores se repitan con mayor frecuencia que otros. Al que aparezca el mayor número de veces, el de mayor frecuencia relativa, se le llamará moda (M_o). Si es uno sólo de esos valores el que más se repite, el estudio será unimodal. Si son dos, bimodal, y, si son tres, trimodal.

En el caso de magnitudes continuas el intervalo modal será el que contenga la moda.

4. 1. 3 Mediana

Una vez obtenidos los datos de la muestra observada, si se ordenan siguiendo un determinado criterio de crecimiento o decrecimiento, se llama mediana (M_e) al valor del dato que ocupa el lugar central, si el número de datos es impar, o bien la media aritmética de los dos valores centrales en el caso de que el número de datos sea par.

Si la variable analizada es cuantitativa pero continua (estudiada por medio de intervalos), el valor de la mediana se calculará de la siguiente manera:

$$M_e = a + x$$

siendo:

$$x = \frac{b-a}{n_i}\left(\frac{N}{2} - \sum_{j \prec i} n_j\right)$$

donde $[a, b]$ es el intervalo mediano (intervalo que contiene el valor de la mediana) y

$$\sum n_j \left(j = 1,2,\ldots\ldots,i-1\right)$$

la frecuencia absoluta acumulada hasta el intervalo $[a, b]$, excluido este.

4. 1. 4 Cuantiles

Son parámetros estadísticos que consideran a la población dividida en un número de partes iguales, con el fin de situar determinados valores por encima o por debajo de uno prefijado.

Si las partes en que se divide la población son 100, los parámetros se llamarán *centiles* o *percentiles* y serán 99.

Si se divide en 10 partes iguales, se llamarán *deciles* y serán 9.

Si se divide en 4 partes iguales, *cuartiles* y serán 3. El segundo cuartil coincidirá con la mediana y dividirá a la población en dos partes iguales.

El cálculo de los distintos tipos de Cuantiles se resolverá de forma similar al utilizado para la mediana.

Ejemplo 7.

Al analizar los resultados obtenidos en un examen de Geometría de 1º de Arquitectura y tras llevar a cabo un estudio de las notas conseguidas por los alumnos, un profesor informa a un estudiante que ha aprobado con un 7 y que se encuentra en el decil 6 (D_6) respecto del resto de la clase.

En estas circunstancias, el alumno dejaría al 60% de sus compañeros por debajo de su puntuación o, lo que sería equivalente, se encontraría entre el 40% de los alumnos con mejores calificaciones de la clase.

4.1.5 Ejercicios resueltos

Ejercicio 1.

Elegida, de forma aleatoria, una muestra de 33 individuos de una población y después de analizar su reacción ante un determinado estímulo, se han tomado las siguientes medidas en segundos**.**

51, 60, 56, 64, 56, 63, 61, 63, 57, 50, 62, 49, 70, 72, 54, 48, 53, 58, 66, 68, 45, 74, 65, 58, 61, 62, 59, 64, 57, 63, 52, 67, 55.

A partir de estos datos se pide calcular los siguientes parámetros:

a) Media
b) Mediana
c) Primer y tercer cuartil

Solución:

a) Para calcular la media sumamos todos los datos y dividimos por el número de ellos. La media vendrá medida en segundos. Sabiendo que $\overline{x}$ viene dada por la expresión:

$$\overline{x} = \frac{x_1 + x_2 + \dots\dots\dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

la media pedida sería

$$\overline{x} = \frac{1963}{33} = 59{,}48$$

b) Para hallar la mediana, en principio habrá que ordenar los valores en orden creciente o decreciente. Lo hacemos en orden creciente.

45, 48, 49, 50, 51, 52, 53, 54, 55, 56, 56, 57, 57, 58, 58, 59, 60, 61, 61, 62, 62, 63, 63, 63, 64, 64, 65, 66, 67, 68, 70, 72, 74.

Al ser el número de elementos impar, el valor que divide la serie en dos partes iguales, el que ocupa el lugar central, el 60, sería la mediana**.**

45, 48, 49, 50, 51, 52, 53, 54, 55, 56, 56, 57, 57, 58, 58, 59, **60**, 61, 61, 62, 62, 63, 63, 63, 64, 64, 65, 66, 67, 68, 70, 72, 74.

c) También una vez ordenados, el primer cuartil dejará hacia la derecha tres cuartas partes de la población, el 75%, de los valores de la muestra; y hacia la izquierda una cuarta parte, el 25%. El valor que cumple la condición es el 55. El tercer cuartil, el valor que deje hacia la derecha el 25% de los datos y hacia la izquierda el 75%, será el 64.

45, 48, 49, 50, 51, 52, 53, 54, **55**, 56, 56, 57, 57, 58, 58, 59, 60, 61, 61, 62, 62, 63, 63, 63, **64**, 64, 65, 66, 67, 68, 70, 72, 74.

El número de datos que se encuentra entre el primer y tercer cuartil representa el *recorrido intercuartílico.*

Ejercicio 2.

Tomando como datos los valores facilitados en el ejercicio 1 se pide:

a) Elaborar unas tablas estadísticas que agrupen los datos por intervalos.
b) Volver a calcular la Media y la Mediana en las nuevas condiciones comparando los resultados con los obtenidos en el ejercicio anterior.

Solución:

La variable se va a estudiar como continua, agrupando los datos en 5 intervalos de igual amplitud, y se va a volver a calcular la media y la mediana.

En la tabla 5 se han representado:

Columna 1: Agrupación de los datos obtenidos en intervalos de amplitud 6 unidades.

Columna 2: Frecuencias absolutas (número de datos incluidos en ese intervalo). La suma de frecuencias absolutas supone el total de la población.

Columna 3: Marcas de clase de cada intervalo (media aritmética de los valores numéricos de sus extremos).

Para el cálculo de la media aritmética se utilizan ahora las marcas de clase (m_i) de cada intervalo, en lugar de los valores (x_i) que tomaba la variable, multiplicadas por las frecuencias de su intervalo correspondiente.

Tiempos	Frecuencias absolutas n_i	Marcas de clase m_i	$m_i n_i$	Frecuencias absolutas acumuladas N_i
$[45-51)$	4	48	192	4
$[51-57)$	6	54	324	10
$[57-63)$	11	60	660	21
$[63-69)$	9	66	594	30
$[69-75)$	3	72	216	33

Tabla 5

Como en este caso

$$\bar{x} = \frac{\sum m_i n_i}{\sum n_i}$$

obtenemos para la media

$$\bar{x} = \frac{1986}{33} = 60{,}18$$

y para la mediana

$$M_e = 57 + \frac{(16{,}5 - 10).6}{11} = 60{,}5$$

Al comparar los parámetros calculados en las nuevas condiciones se observan diferencias con los obtenidos anteriormente. Se deben a que el hecho de agrupar datos supone que sus diferentes valores se distribuyen uniformemente dentro de los intervalos. Si eso fuera cierto, que no lo es, los valores coincidirían.

Ejercicio 3.

Elegida una muestra aleatoria de 160 sujetos se ha observado el grado de aptitud para desempeñar su cometido. Como consecuencia, se han obtenido una serie de mediciones que han sido agrupadas, por intervalos de amplitud uniforme de 10 unidades, para elaborar la siguiente tabla de frecuencias (Tabla 6).

X_i	n_i
$[0-10)$	8
$[10-20)$	22
$[20-30)$	32
$[30-40)$	44
$[40-50)$	28
$[50-60)$	20
$[60-70)$	6
	Total 160

Tabla 6

Utilizando estos datos de pide calcular y comentar los siguientes puntos:

a) ¿Entre qué valores se encuentra el 50% central de los sujetos?
b) Calcular el percentil 27.
c) ¿A partir de que puntuación se encuentra el 12 % de los sujetos más aptos?
d) Si no se tienen en cuenta ni el 15% de los sujetos menos aptos y el mismo porcentaje de los más aptos ¿En qué intervalo se encontrarían el resto de sujetos de la muestra?

Solución:

Vamos a ir completando la tabla facilitada con las columnas necesarias para calcular y comentar los distintos apartados que se piden.

X_i	n_i	N_i
$[0-10)$	8	8
$[10-20)$	22	30
$[20-30)$	32	62
$[30-40)$	44	106
$[40-50)$	28	134
$[50-60)$	20	154
$[60-70)$	6	160

Tabla 7

La tercera columna representa las frecuencias absolutas acumuladas. El último término de la columna coincidirá con el tamaño de la muestra de partida: 160 individuos (Tabla 7).

a)
Tendremos que calcular el primer y tercer cuartil que dejarán el 25% de la muestra por debajo y por encima respectivamente. Entre ambos quedará el 50% restante. Al ser 160 los elementos de la muestra la cuarta parte serán 40. La primera frecuencia acumulada que supera este valor es 62 que pertenece al tercer intervalo $\left[20-30\right)$. Por lo tanto, el primer cuartil será:

$$Q_1 = 20 + \frac{(40-30)10}{32} = 23{,}125$$

Las tres cuartas partes de la muestra contienen 120 elementos. El tercer cuartil estará, por lo tanto en el quinto intervalo $\left[40-50\right)$ y su valor será:

$$Q_3 = 40 + \frac{(120-106)10}{28} = 45$$

b)
El 27% de la muestra es 43,2. Por lo tanto, el percentil 27 estará en el tercer intervalo $\left[20-30\right)$. Su valor será:

$$C_{27} = 20 + \frac{(43{,}2-30)10}{32} = 24{,}125$$

c)
El valor que deja por encima el 12% de los individuos más aptos es el mismo que deja por debajo al 88% de los menos aptos, y coincide con el percentil 88. El 88% de la muestra resulta 140,8.

$$C_{88} = 50 + \frac{(140{,}8-134)10}{20} = 53{,}4$$

d)
Tenemos que calcular el percentil 15 y el percentil 85. El 15% del tamaño de la muestra es 24 y el 85% es 136. Aplicando la fórmula correspondiente a los percentiles en sus respectivos intervalos resulta:

Para el percentil 15

$$C_{15} = 10 + \frac{(24-8)10}{22} = 17,27$$

y para el percentil 85

$$C_{85} = 50 + \frac{(136-134)10}{20} = 51$$

El intervalo comprendido entre esos dos parámetros incluiría el 70% de los valores centrales.

4.2 Parámetros estadísticos de dispersión

Los parámetro estudiados hasta ahora, los parámetros de posición, los que hacen referencia fundamentalmente a la forma de reparto de las frecuencias de los datos obtenidos de la observación del experimento analizado, venían representados por valores que pertenecían al intervalo en el que se estudiaba la variable.

Los parámetros de dispersión, en cambio, atienden especialmente a la separación de los distintos elementos de la distribución, y son: la *desviación media*, la *desviación típica* y la *varianza*.

4.2.1 Desviación media

Para definir este primer parámetro vamos a suponer que una vez analizado un suceso aleatorio S obtenemos de su observación los valores:

$$x_1, x_2, ..., x_m$$

De frecuencias absolutas respectivas:

$$n_1, n_2, ..., n_m$$

Se define *desviación media* ($D_{\bar{x}}$) como la media aritmética de las distancias de los valores a la media, es decir:

$$D_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}| n_i$$

4.2.2 Desviación típica

Siguiendo con el supuesto anterior, se define la *desviación típica* (σ) de los valores x_i obtenidos de una determinada variable como:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{n} n_i}}$$

que equivale a:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \bar{x}^2} \ .$$

4.2.3 Varianza

Se denomina *varianza* al cuadrado de la desviación típica y se representa por σ^2.

La desviación típica y su cuadrado, la varianza, son parámetros estadísticos que complementan a otros, como por ejemplo la media, al ser indicativos de la cercanía o alejamiento de los valores obtenidos a uno marcado previamente.

4.2.4 Ejercicios resueltos

Ejercicio 4.

Se ha medido una determinada variable en un conjunto de individuos obteniéndose las siguientes mediciones para esa variable:

3	5	3	6	3	5	5	4	9	8	5	7	3	8	2

Se pide calcular:

a) La Media.
b) La Desviación Típica.

Solución:

a) Si $\bar{x} = \dfrac{x_1 + x_2 + ... + x_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, entonces $\bar{x} = 5$

b) Si $\sigma = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \bar{x}^2}$, entonces $\sigma = S = 2{,}03$

Tanto la media aritmética como la desviación típica se miden en las mismas unidades que la variable analizada.

Cabe la posibilidad, y eso es lo más frecuente hoy día, de que el número de datos que se nos faciliten para trabajar con ellos en un estudio estadístico sea muy alto.

En ese caso, los cálculos y los posteriores trabajos a llevar a cabo: tablas, gráficos y análisis, serían muy engorrosos como para hacerlos a mano o con una simple calculadora.

Hoy día el trabajo se ha facilitado enormemente con la ayuda de los ordenadores y con la existencia de numerosos programas informáticos que no solo ayudan en el cálculo, sino que están expresamente diseñados para el tratamiento estadístico de un conjunto de datos por muy elevado que éste sea.

El ejercicio que se plantea a continuación (Ejercicio 5) contiene dos ejemplos de un mismo enunciado. En él se facilitará una serie de datos numéricos con los que se pide trabajar construyendo tablas de frecuencias y obteniendo parámetros estadísticos.

Ejercicio 5.

En el primero (Ejercicio 5.1) se facilitan, con el conjunto de datos, las tabla de frecuencias y diversos parámetros estadísticos calculados.

En el caso de que, en un determinada situación, tengamos que elaborar un informe de tipo más o menos general, contestar a preguntas concretas, sacar conclusiones, intentar predecir comportamientos futuros.... utilizando tablas similares a las del ejercicio 5. 1, en las que figuran una gran cantidad de datos; habrá que seleccionar los que realmente sirvan a nuestro fin dejando de lado aquellos otros que, aún aportando una información coherente, real y positiva, no sirva a nuestros fines.

Una vez elaboradas las tablas y calculado los parámetros, elegidos y analizados aquellos que se consideren oportunos, siempre es útil redactar las conclusiones pertinentes.

Ejercicio 5.1.

(Ejemplo a seguir en el ejercicio 5.2)

14	17	16	14	15
16	15	22	19	16
19	20	22	18	17
20	20	18	19	14
14	22	16	20	14
18	21	20	21	23
18	14	21	14	15
23	19	15	14	19
22	21	22	19	18
21	17	14	16	18
15	23	21	21	22
23	23	17	23	19
18	14	15	18	14
22	23	16	23	21
14	14	21	15	19
15	15	21	17	16
21	21	21	14	16
17	18	22	19	15
19	17	20	15	17

x_i	n_i	n_a	$f_i n_i$	$n_i x^2$
14	14	14	196	2744
15	11	25	165	2475
16	8	33	128	2048
17	8	41	136	2312
18	9	50	162	2916
19	10	60	190	3610
20	6	66	120	2400
21	13	79	273	5733
22	8	87	176	3872
23	8	95	184	4232
			1730	**32342**

N = 95
$\sum(x_i) = 1730$
$\sum(x^2) = 32342$
Moda = 14
Mediana = 18

Mínimo = 14
Máximo = 23
Rango = 9
$\sigma = 2{,}9696526$
$\sigma^2 = 8{,}81883657$
$s = 2{,}98540684$
$s^2 = 8{,}91265398$

Ejercicio 5.2.

En este segundo ejemplo sólo se dan los datos con los que se deben elaborar tablas y calcular parámetros como en el ejercicio anterior.

13	22	19	18	19
17	13	13	15	17
18	20	13	14	18
15	18	21	18	17
19	21	17	15	13
17	17	21	15	16
17	18	13	21	20
13	22	17	22	21
16	19	14	14	22
18	15	15	13	17
14	17	15	20	22
16	21	13	18	14
22	21	16	14	13
22	20	13	20	16
21	18	20	21	15
21	14	14	18	17
16	15	21	20	20
18	17	19	22	20
22	16	18	17	15

5 Bibliografía

1. G. Garbacho, A. *Estadística elemental moderna*. Ed. Ariel, Barcelona, 1987. ISBN: 84-344-2005-8

2. Turner, J.C. *Matemática moderna aplicada. Probabilidades, estadística e investigación operativa*. Alianza Editorial, Madrid 1986. ISBN: 84-206-2092-0

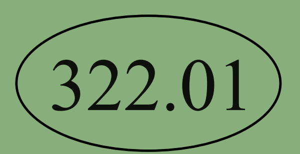
3. Castillo Ballesteros, V. del; Ortega Vidal, C.; Ruiz Domínguez, M.; Sánchez González, J. M. *Matemáticas 2* . Editorial Bruño. Madrid, 1989. ISBN: 84-216-1101-1

4. Castillo Ballesteros, V. del; Ortega Vidal, C. Ruiz Domínguez, M.; Sánchez González, J. M. *Matemáticas 2* . *Problemas resueltos.* Editorial Bruño. Madrid, 1990. ISBN: 84-216-1262-X

5. Colera, J.; Oliveira, M.; J. García, R. *Matemáticas Aplicadas a las Ciencias Sociales II*. Anaya. Barcelona, 2001. ISBN: 84-667-0125-7

6. Colera, J.; Oliveira, M. J.; García,; R. Fernández, S. *Matemáticas Aplicadas a las Ciencias Sociales I*. Anaya. Barcelona, 2001. ISBN: 84-207-9911-4

CUADERNO



Cuadernos.ijh@gmail.com
info@mairea-libros.com

